

文章编号: 1005-0523(2008)05-0077-04

聚类分析在开放式基金绩效研究中的应用

闫云娟, 冯大一

(华东交通大学 基础科学学院 江西 南昌 330013)

摘要: 聚类分析在基金绩效的研究方面有很大的挖掘空间. 为研究基金的绩效, 从基金的规模发展、收益如何及成长性 3 个方面进行了考察, 选取了 2008 年 1 月到 2008 年 6 月共 24 周的数据, 其中任意选取 30 只基金, 以其周收益率为基本数据计算并建立了 6 个指标. 通过 Q 聚类分析方法, 分析了这 30 只基金在整个股市大幅下跌的表现情况. 以此来判断基金的绩效, 来帮助投资者做出较为准确的投资.

关键词: 开放式基金; 聚类分析; 系统聚类法

中图分类号: O212.4; F830.59

文献标识码: A

1 研究背景

截止到 2008 年 6 月, 我国基金已经发展到 311 支. 随着我国证券市场如火如荼的快速发展, 基金投资已成为许多人的投资手段, 如何抓住机会, 规避风险已成为众多投资者所关心的问题. 除了关注各项政策等定性因素以外, 我们还可以利用基金过去已有的信息来得到一些定量的信息.

目前对基金业绩的考察可以通过直接比较, 即将基金与某个标准来进行比较, 其优点是一目了然但不能充分利用数据得到其内在联系; 另外还可以间接比较, 就是将基金与该基金相似的一组进行比较, 对同类型的基金业绩进行比较分析来评定其业绩的优劣. 本文利用系统聚类分析的方法期望得到业绩突出, 收益良好的投资基金.

2 研究方法

传统方法通常应用多个指标来定性分析业绩. 聚类分析就是分析如何对样品(或变量)进行量化分类的方法. 通常聚类分析分为 Q 型聚类和 R 型聚类^[1]. Q 型是对样品进行分类处理, R 型是对变量进行分类处理. 本文应用 Q 型聚类的方法来处理数据, 其优点就是能够应用多个变量对样本进行分类. 聚类分析所得到的结果比传统方法得到的结果更客观、全面、细致.

2.1 样品相似性的度量

距离作为对样品之间的相似程度的度量是聚类分析的基础. 样品的相似程度可用两点间距离公式来度量. 从不同角度定义有以下三种:

设每个样品有 p 个指标从不同方面描述其性质, 形成一个 p 维向量.

(1) Minkowski 距离

$$d_{ij}(q) = \left(\sum_{k=1}^p |x_{ik} - x_{jk}|^q \right)^{1/q}$$

收稿日期: 2008-07-25

作者简介: 闫云娟(1979-), 女, 山西阳泉人, 讲师, 硕士, 研究方向为应用数学.

$$\text{当 } q = \begin{cases} 1 & \text{绝对距离} \\ 2 & \text{欧几里德距离} \\ \infty & \text{切比雪夫距离} \end{cases}$$

$$(2) \text{ Mahalanobis 距离} \quad d_{ij}^2(M) = (X_i - X_j)' \Sigma^{-1} (X_i - X_j)$$

X_i 与 X_j 是来自均值向量为 μ , 协方差为 Σ 的总体 G 中的 p 维样品. 马氏距离又称为广义欧几里德距离, 它考虑了观测变量的关系.

$$(3) \text{ 兰氏距离} \quad d_{ij}(L) = \frac{1}{p} \sum_{k=1}^p \left| \frac{X_{ik} - X_{jk}}{X_{ik} + X_{jk}} \right|$$

本文采用的是欧氏距离, 即欧几里德距离来度量样品间的距离.

2.2 类间距离

在系统聚类之前, 我们首先要定义类间距离, 由类间距离的不同定义产生了不同的系统聚类法. 常用的类间距离有 8 种之多, 与之相应的系统聚类法也有 8 种. 本文只介绍 3 种比较常用的方法, 分别为:

$$(1) \text{ 最短距离法} \quad D_{kr} = \min_{X_i \in G_i, X_j \in G_j} d_{ij} = \min \left\{ \min_{X_i \in G_k, X_j \in G_p} d_{ij}, \min_{X_i \in G_k, X_j \in G_q} d_{ij} \right\}$$

设类 G_p 与 G_q 合并成新的类, 记为 G_r . D_{kr} 表示任意一类 G_k 与 G_r 的距离.

$$(2) \text{ 最长距离法} \quad D_{kr} = \max_{X_i \in G_i, X_j \in G_j} d_{ij} = \max \left\{ \max_{X_i \in G_k, X_j \in G_p} d_{ij}, \max_{X_i \in G_k, X_j \in G_q} d_{ij} \right\}$$

$$(3) \text{ 类平均距离法} \quad D_{kr} = \frac{1}{n_p n_q} \sum_{i \in G_p} \sum_{j \in G_q} d_{ij}$$

即用 G_p, G_q 每两两样品间距离的平均值作为这两类之间的距离.

2.3 系统聚类法

基本思想: 距离相近的样品先聚类, 距离相远的样品后聚类, 过程一直进行下去, 每个样品总能聚到合适的类里去. 系统聚类过程是: 假设共有 n 个样品, 第一步将每个样品聚成一类, 共有 n 类; 第二步根据所确定的样本距离, 把距离较近的两类样品聚为一类, 其他样品仍自聚一类, 共聚 $n-1$ 类; 第三步将距离最近的两个类进一步聚成一类, 共聚 $n-2$ 类; 以上步骤一直进行下去, 最后将所有的样品都聚成一类. 为了直观的反映系统聚类过程, 通常会画出其聚类谱系图. 本文采用类平均距离法进行系统聚类.

3 实证研究

3.1 数据的选取

本文共选取了 2008 年 1 月到 2008 年 6 月共 24 周的数据, 其中任意选取 30 只基金, 以其周收益率为基本数据而计算并建立了以下一系列指标. 原始数据由中国证券报与和讯基金网获得. 由于篇幅有限且原始数据量比较大, 文中只给出了 6 个指标数据和聚类的结果.

3.2 聚类分析指标体系的选取^[2]

X_1 为上市时间, X_2 为夏普指数, X_3 为夏普指数, X_4 为特雷诺指数, X_5 为单位净值, X_6 为基金变动率^[3], 由于 2008 年只统计出第 1 季度的基金规模, 所以本文采用 2008 第 1 季度与 2007 年全年的基金规模做比较.

$$(1) \text{ 周收益率计算方法: } r_{it} = (\text{nav}_{it} - \text{nav}_{i,t-1}) / \text{nav}_{i,t-1}$$

其中: r_{it} 表示第 i 只基金在第 t 周的周收益率; nav_{it} 表示第 i 只基金在第 t 周周末的单位净值; $\text{nav}_{i,t-1}$ 表示第 i 只基金在第 $t-1$ 周周末的单位净值.

(2) 詹森指数^[4]: 它是对基金超额收益的一种衡量指标, 其越大, 则越能体现基金经理的才能, 基金越优.

$$J = E(r_{it}) - [r_f - E(r_{mt} - r_f)] \beta_i$$

$$\beta_i = \frac{r_p - r_f}{r_m - r_f}$$

其中: $E(r_{it})$ 为第 i 只基金在整个研究期间的平均周收益率; β_i 为系统风险; r_p 表示基金的平均周投资收益率; r_f 为周无风险收益率, 本文以 2008 年 1 年的定期存款利率 4.14% 为这 24 周的无风险利率, 所以周无风险利率为 0.086 25%; r_m 为市场组合周收益率; r_{mi} 为市场基准周收益率。

(3) 特雷诺指数: $T = (r_p - r_f) / \beta$, 它表示投资组合平均收益率超出无风险收益率部分(超额收益率)与投资组合的净值之商, 即基金承担单位系统风险所获得的超额收益^[5]。

(4) 夏普指数: $s_p = (r_p - r_f) / \sigma$, 其中 σ 表示基金周收益率的标准差。夏普指数实际上是投资组合平均周收益率超出周无风险收益率部分(超额收益率)与投资组合收益率的标准差之商, 即基金承担单位风险(包括系统风险和非系统风险)所获得的超额收益。

指标数据见表 1。

表 1 聚类指标表

基金名称	X_1 /月	X_2	X_3	X_4	X_5 /元	X_6
鹏华动力增长	15	-0.050 6	0.001 6	-0.003 1	1.279	15.79
大成沪深 300	26	-0.087 2	-0.000 5	-0.004 7	1.002 8	-3.5
50ETF	40	-0.067 8	0.000 6	-0.003 7	2.377	58.04
嘉实 300	33	-0.095 7	-0.001	-0.005 2	0.767	3.82
万家 180	62	-0.070 4	0.000 6	-0.003 8	0.695 6	17.64
广发大盘	10	-0.110 2	-0.000 5	-0.006 5	0.766 2	-11.64
华夏大盘精选	45	0.037 7	0.006 1	0.002 4	5.433	-1.03
上投摩根内需动力	12	-0.013 9	0.003 7	-0.000 8	0.968 2	3.71
华夏优势增长	18	0.023 5	0.005 3	0.001 4	1.873	-3.65
易方达积极成长	45	-0.036	0.002 7	-0.002 1	1.109 1	-10.99
交银成长	18	-0.012 9	0.003 8	-0.000 8	2.008 8	-0.27
上投摩根阿尔法	41	0.004 2	0.004 5	0.000 3	4.533 2	2.53
华夏红利	34	0.044 9	0.006 2	0.002 6	2.406	31.71
中海优质成长	53	-0.078 6	0.001	-0.004 7	0.647 2	-5.54
红利 ETF	17	-0.076 2	-0.000 7	-0.004 3	2.501	23.93
深 100ETF	26	-0.073 6	-0.000 1	-0.004 2	3.116	2.45
180ETF	25	0.078 1	-0.000 1	-0.004 2	6.718	21.27
中银收益	29	-0.095 7	0.000 6	-0.005 6	0.758 2	-10.01
交银蓝筹	9	-0.159 4	0.000 5	-0.009	0.747 6	-10.01
泰达荷银	38	0.037 8	0.005 2	0.002 3	1.286 1	6.85
华安宝利配置	45	-0.032 8	0.003 2	-0.001 9	0.843	70.38
华宝兴业先进	18	-0.063 5	0.001 7	-0.003 7	2.008 2	-5.93
金鹰中小盘精选	48	-0.079 8	0.000 7	-0.005 2	1.176 9	165.85
招商债券基金	60	-0.086 7	0.004 1	-0.011	1.138 7	341.52
南方多利	27	0.089 9	0.004 8	0.007 7	1.071 6	13.28
华夏回报	56	0.050 5	0.005 7	0.003	1.152	-8.89
华安创新	78	-0.068 3	0.001 8	-0.004	0.697	-5.26
嘉实理财增长	59	-0.063 4	0.002	-0.003 9	3.329	14.69
南方避险增值	49	-0.061 1	0.002 8	-0.003 7	2.105 2	-3.37
长城久泰	47	-0.091 9	-0.005	-0.007	1.135 1	-7.82

3.2 聚类分析结果

在进行系统聚类之前, 为消除量纲的影响, 首先将数据标准化处理。根据 SPSS^[6] 软件计算得聚类谱系图(见图 1)。

从中我们发现大成沪深、嘉实 300 距离最近, 首先聚为一类, 随后以中银收益、鹏华动力、华宝兴业、红利

ETF、深100ETF、广发大盘、交银蓝筹、万家180、中海优质、华安创新、易方达积、南方避险的顺序聚为第一类。这类基金上市时间较长,规模变化不大,就基金经理在应对这半年来股市大幅下跌显得比较有经验,有一定的避险能力,尽管随着沪证和深证股指的整体下滑,这些基金的回报率也在下滑,但基金管理人的选股及结构投资能力仍值得我们看好,如果市场重新活跃,该基金是一个不错的选择。

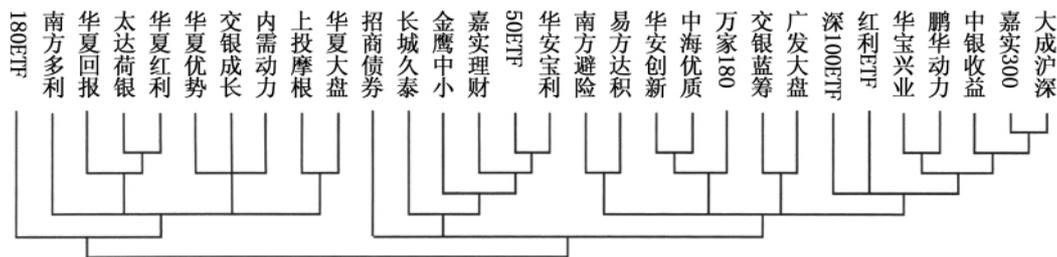


图1 聚类谱系图

第二类是以如下顺序聚为一类的: 华夏大盘、上投摩根阿尔法、上投摩根内需动力、交银成长、华夏优势、华夏红利、泰达荷银、华夏回报、南方多利、180 ETF。

这类基金上市时间较长,同时我们可以看出在最近半年如此低靡的状态下,他们抵御风险的能力还是不错的,因此我们还是可以考虑做长期投资。

第三类是华安宝利、50 ETF、嘉实理财、金鹰中小、长城久泰顺、招商债券这类基金规模变化较大,回报率不高,在这段时间内业绩表现不好。

参考文献:

- [1] 朱建平. 应用多元统计分析[M]. 北京: 科学出版社, 2006. 62-90.
- [2] 张海燕. 利用回归分析对基金业绩进行评价[J]. 中国民航学报, 2003, 21(7): 64-65.
- [3] 李凯, 史金艳, 武珊. 基于聚类分析的封闭式基金折价问题实证研究[J]. 预测, 2006, 25(2): 60-61.
- [4] 胡猛, 向嘉华. 证券投资基金务实[M]. 北京: 社会科学文献出版社, 1999. 110-130.
- [5] 王敬, 王颖. 基于主成分分析的基金绩效评价模型研究[J]. 大连理工大学学报, 2005, 46(2): 72-77.
- [6] 章文波, 陈红艳. 实用数据统计分析及SPSS12应用[M]. 北京: 人民邮电出版社, 2006. 178-200.

Application of Cluster Analysis in the Performance of Open-end Funds

YAN Yun-juan, FENG Da-yi

(School of Basic Sciences, East China Jiaotong University, Nanchang 330013, China)

Abstract: There is potential in the research of fund achievement by means of cluster analysis. The aspects including development, income and expanding of fund are inspected. The paper selects 30 funds randomly in 24 week-long data from January 2008 to a June 2008. Six targets are established and calculated based on the week returns ratio as the basic data. The performance of these 30 funds is analyzed in the situation which falls largely in the entire stock market. Judging the fund by the achievement may help the investors to make a more accurate investment.

Key words: open-end fund; cluster analysis; hierarchical cluster analysis

(责任编辑: 刘棉玲)